

**WE CLAIM:**

1. A system for serving web pages to a client in response to a client request specifying a resource, the system comprising:

5 a first web server having a first address, the first web server having a request interface for receiving requests and a response interface for sending responses to the requests;

10 a communication network coupled to the first web server;

15 a second web server having a second address and implementing interface processes for communicating with a client application and for communicating with the first web server, the second web server coupled to both obtain at least some of the request-specified resources from the first web server and augment the obtained resources using resources within the second web server to generate a response to the client request.

2. The system of claim 1 wherein the first web server is located at a web site hosting center and implements multiple web sites.

3. The system of claim 1 wherein the second web server further comprises:

means for detecting a busy condition in the first web server; and

5 means for substituting resources within the second web server for the resources within the first web server when the busy condition is detected.

4. The system of claim 3 wherein the substituted resources comprise a redeemable coupon.

5. The system of claim 1 wherein the second web server further comprises:

means for detecting a condition in which the first web server is unavailable; and

- 5 means for substituting resources within the second web server for the specified resources when the first web server is unavailable.

6. The system of claim 5 wherein the substituted resources comprise a redeemable coupon.

7. A method for serving web pages to a client in response to a client request specifying a network resource, the method comprising:

- providing an originating web server upon which the  
5 requested resource resides, the originating web server having a request interface for receiving requests and a response interface for sending responses to the requests;  
providing a front-end server;  
redirecting a client request addressed to the  
10 originating web server to the front-end server;  
using the front-end server to obtain at least some of the request-specified resources from the originating web server;  
augmenting the obtained resources using resources  
15 within the front-end server to generate a response to the client request; and  
communicating the augmented response from the front-end server to the client.

8. A web site comprising:

- an originating web server coupled to the network and executing software to access a first set of content and functionality in response to requests from a web client;  
5 a front-end server coupled to the network and executing software to access a second set of content and functionality in response to requests from the web client;

10 a first communication channel supporting request and response communication between the web client and the front-end server;

a second communication channel supporting communication between the front-end server and the originating web server; and

15 means operating cooperatively between the originating web server and the front-end server for receiving requests for web pages from the web client and serving web pages to the requesting web client in response to the received requests.

9. The web site of claim 8 further comprising means for directing request and response traffic from the first channel to the second channel.

10. A system for serving content from a plurality of network resources comprising:

a first server that is an originating web server for a first set of resources;

5 a second server that is an originating web server for a second set of resources;

a request interface within the first server for receiving requests specifying the first server, the first server responding to the request by generating a response using the first set of resources and generating a request to the second server in order to use the second set of resources when generating the response; and

10 a request interface within the second server for receiving requests specifying the second server, the second server responding to the request by generating a response using the second set of resources and generating a request to the first server in order to use the second set of resources when generating the response.

11. The system of claim 10 wherein the first and second set are not mutually exclusive.

12. The system of claim 10 wherein the first and second sets are mutually exclusive.

13. A method for serving content from a plurality of servers comprising:

generating a request for content, the request being associated with a specified address;

5 providing a server having the address specified in the request;

directing the request to the server having the specified address and having at least a portion of the network resources specified in the request;

10 causing the server to serve a first portion of the request from the server itself; and

causing the server to serve a second portion of the request from a second tier web server.

14. A system for serving content from a plurality of network resources comprising:

a first web server comprising a first set of resources and coupled to serve requests received from a first plurality of network appliances by accessing the first set of resources in response to requests generated by the first plurality of network appliance;

10 a second web server comprising a second set of resources and coupled to serve requests from a second plurality of network appliances by accessing the second set of resources in response to requests generated by the second plurality of network appliances; and

a communication channel between the first and second web servers such that a request received on the first web server for content of the second web server are served by

causing the first web server to access the second web server to obtain the content and servicing the request from the first web server.

15. The system of claim 14 wherein requests received on the second web server for content of the first web server are served by causing the second web server to access the first web server to obtain the content and servicing the request from the second web server to obtain content from content source, the originating web server having a request interface for receiving requests from a network and a response interface for sending responses to the requests.

16. The system of claim 14 wherein the first tier web server fetches content from the second tier web server and serves the fetched content such that it is the originating server for the fetched content.

17. A system for serving web pages to a network client comprising:

a network;

an origin web server coupled to the network;

5 a client coupled to the network and generating requests directed to the origin web server;

a gateway machine in communication with the client, the gateway machine configured to receive requests from the client, obtain web pages from the origin web server in response to the received request, and serve the obtained pages to the client;

10 means for detecting a busy condition in the origin server;

means responsive to the busy condition for serving a page generated by the gateway machine in response to the client request.

18. The system of claim 17 wherein the gateway machine comprises web server software capable of generating static and dynamic web pages.

19. The system of claim 17 wherein the means for serving substitute pages places the gateway machine in a role of the origin server.

20. A method for serving web pages to a network client comprising:

generating requests from the network client, the requests specifying a network resource that is configured to generate a desired web page;

routing the requests through a gateway machine that maintains a connection with the network resource;

detecting a busy condition in the network resource; and

serving an alternate page from the gateway machine to the network client in response to detecting a busy condition.

21. A system for serving web pages to a network client comprising:

a network;

an origin web server coupled to the network;

a client coupled to the network;

a gateway machine in communication with the client, the gateway machine configured to receive requests from the client, obtain web pages from the origin web server in response to the received request, and serve the obtained pages to the client;

means for detecting when the origin server becomes unavailable; and

means responsive to origin server becoming  
unavailable for serving an alternate page from the  
15 gateway machine to the client.

22. A system for caching network resources  
comprising:

an origin web server operable to provide responses  
containing network resources in response to requests for  
5 the contained resources;

an first intermediary server having a cache therein  
and coupled to receive the responses and to receive the  
requests for network resources;

mechanisms within the first intermediary server for  
10 determining whether the requested network resources are  
within the cache; and

mechanisms within the first intermediary server for  
serving the requested network resources from the cache  
when requested network resources are within the cache.

23. The system of claim 22 wherein the network  
resources are associated with a priority value and the  
cache comprises a priority sensitive cache that  
selectively stores the network resources in the cache  
5 based at least in part on the associated priority.

24. The system of claim 22 further comprising a  
second intermediary server having a cache therein, the  
second intermediary server being coupled to coupled to  
receive the responses and to receive the requests for  
5 network resources.

25. The system of claim 24 further comprising a  
communication channel coupling the first intermediary  
server and the second intermediary server.

26. The system of claim 25 wherein the network resources are associated with a priority value and the communication channel is responsive to the priority value to regulate the manner in which data is communicated  
5 between the first intermediary server and the second intermediary server.

27. The system of claim 22 wherein the cache comprises a passive cache that caches network resources as they pass through the intermediary server.

28. The system of claim 22 wherein the cache comprises an active cache that caches network resources in response to encoded instructions within the network resource.

29. The system of claim 22 wherein the cache comprises an active cache that caches network resources in anticipation of requests for the network resource.

30. The system of claim 22 wherein the cache comprises an active cache that caches network resources in response to instructions from another intermediary server.

31. The system of claim 22 wherein the intermediary server comprises a front-end server located on a client-side of a network.

32. The system of claim 22 wherein the intermediary server comprises a back-end server located on an originating server-side of a network.

33. The system of claim 22 wherein the intermediary server comprises:

a front-end computer having a cache therein and located on a client-side of a network; and



5 a back-end server having a cache therein located on  
an originating server-side of a network.

34. A method for serving web pages in response to  
requests from a network client, wherein the requests  
specify content desired by the user of the network  
client, the method comprising:

5 providing a gateway server configured to receive  
requests specifying content;

providing a plurality of network servers, at least  
one of the network servers housing the specified content,  
and at least one of the network servers housing  
10 alternative content;

in response to a received request, generating  
requests in the gateway server to at least one of the  
network servers;

in response to the request received from the gateway  
15 server, generating requests in the at least one network  
server to at least one other network server;

serving a response from the gateway server, wherein  
the response includes content selected from the group  
consisting of specified content from the gateway server,  
20 specified content from one of the network servers,  
specified content from an origin server, alternate  
content from the gateway server, and alternate content  
from one of the network servers.

35. The method of claim 34 further comprising the  
act of routing the requests to a gateway server.

36. The method of claim 34 wherein the selection of  
content included in the response from the gateway server  
is selected by the gateway server.

